DOCUMENT RESUME

ED 050 411 CG 006 416

AUTHOR Thornton, George C., III
TITLE The Validity of Assessment Centers.
INSTITUTION Colorado State Univ., Ft. Collins.
PUB DATE May 71
NOTE 36p.; Paper presented at the Rocky Mountain Psychological Association Convention in Denver, Colorado, May 12-15, 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Evaluation, Evaluation Techniques, *Industrial Personnel, *Performance Tests, *Personnel Evaluation, *Validity

ABSTRACT

This paper summarizes a review of the recent literature in search of evidence for the validity of industrial assessment centers. The topic is divided into two parts: (1) the evidence regarding the validity of several of the individual assessment techniques that are used in industrial assessment centers; and (2) the evidence concerning the validity of the overal assessment rating (OAR). Findings indicate that: (1) the most valid techniques seem to be performance exercises involving "samples" of behavior; (2) few studies have treated the technique psychometrically; (3) numerous reports fail to report the correlations of the tests, games, etc. with any sort of criteria, although these data were gathered; (4) the overwhelming finding is that OAR's are related to a variety of criteria for subsequent performance and potential in managerial positions; (5) there are serious questions regarding the quality of research designs and problems with criterion contamination in some studies; and (6) the overall assessment evaluation adds a unique contribution to assessment beyond information obtainable from objective tests. (Author/TA)

R.4PA, May 1971

The Validity of Assessment Centers

George C. Thornton, III
Colorado State University
Fort Collins, Colorado

This paper summarizes a review of the recent literature in search of evidence regarding the validity of industrial assessment centers. The topic has been divided into two parts: (1) the evidence regarding the validity of several of the individual assessment techniques that are used in industrial assessment centers, and (2) the evidence concerning the validity of the overall assessment ratings (OAR). In addition, the paper reports the lack of evidence presented with regard to the effectiveness of assessment centers as an approach to management development. The author recognizes that assessment centers are used for many purposes, but has chosen to limit his attention to the above areas for these seem to be the most important ones and they are the ones about which the most has been written. Data for the paper were gathered by consulting references in the standard published sources, reviewing transcriptions of speeches given at various conferences and conventions, and obtaining in-house materials from companies sponsoring assessment centers. Letters requesting material were sent to all companies known to have assessment centers.

Validity of Individual Assessment Techniques

One of the outstanding features of the assessment center approach to personnel evaluations is the gathering of multiple assessments by multiple assessment techniques. The synthesis of the assessment data by several assessors during the review period may be built on complex

interactions of observations which take place over a period of time. The examination of validity of individual assessment techniques may not adequately reflect the complex nature of multiple assessments. At the same time knowledge of the contribution of each technique is of practical and theoretical value.

Table 1 presents a summary of the relationships of several assessment techniques and internal and external criterion. It is important to note that the individual assessment techniques cover a broad range of devices including objective and subjective tests, individual and group exercises, and games and discussions. It is particularly important to note that many of the devices are performance exercises that simulate major dimensions of managers' jobs. The table does not reveal the wide variety of ingenious exercises which have been devised. Numerous varieties of in-baskets, role-playing exercises, games, etc. have been developed. Unfortanately, little systematic data has been gathered to evaluate their effectiveness.

The criteria used for this first part of the paper include both internal criteria (e.g. ratings by assessment center staff, and overall assessment ratings (OAR)) and external criteria (e.g. current rank in company, subsequent promotion, salary progress, and managers' ratings). Validation against the OAR assumes that this summary is a relevant, worthwhile criterion. It is a composite of evaluations from multiple assessors often making observations from multiple techniques, but to assign the OAR the status of a "criterion" may be premature. On the other hand, in many situations no better evaluations of managerial characteristics exist. More helpful information can be obtained by examining the relationships with external criteria. These data reveal validity against measures independent of assessment.

To adequately review the validity of the individual assessment techniques would require references to a large body of previous literature in which the techniques have been used by themselves for assessment and/or management development. This paper deals only with the data generated directly from assessment centers. Few studies have treated the performance test psychometrically, that is, generated quantitative data directly from participation in the exercises and correlated these with criteria. For the most part, judgments and ratings are made after the assessors have observed the participants carry out the exercises.

In-basket

An in-basket is a managerial job simulation in which the participant must deal with memos, letters, manuscripts, messages, reports, etc. that commonly would be found in a manager's in-basket. The manager is given appropriate background information about the simulated organization he is to work for, and the job he is to perform. The instructions usually emphasize that the manager is not to play a role or say what he would do with the problems presented to him, but rather that he is to perform as if he, himself, were actually on the job. Each item of the in-basket is a document or set of documents that deals with a particular administrative problem. The responses of the manager are his work on the problem posed by the items; this may take the form of written notes, memoranda, letters, etc. Following the managers' responses, he is usually asked to explain the reasons for his actions.

Bray & Grant (1966) asked assessors to read a narrative report of the assessee's performance in the in-basket itself and in the discussion of reasons for action, and to rate overall in-basket performance on a 5-point scale. Mean ratings were correlated with assessment staff evaluations and subsequent criteria. The in-basket contributed

a sizeable amount to the OAR for both college and no-college groups. Against an external criterion of salary progress the in-basket correlates significantly for 5 of 7 sub-groups. Wollowick & McNamara (1969) found a correlation of 32 between judged in-basket performance and a criterion change of position level over a three-year period.

Leaderless Group Discussion.

In a leaderless group discussion, a group of participants is given a problem to solve and instructed to arrive at a group decision within a specified time period. Often participants are assigned points of view or roles to play. For example they may be instructed to try to get as much salary increase for their subordinate or budget for their department, but at the same time facilitate a harmonious decision. Assessors observe which participants assume various leadership roles, how individuals are accepted by the group, and the influence each participant has on the final decision.

For the college and non-college groups in the Management Progress Study, Bray & Grant (1966) found that judgments of assessors observing the leaderless group discussion yielded the most consistently significant relationships with OAR and salary progress. Wollowick & McNamara (1969) also found a significant correlation with the longitudinal criterion.

Management Games

Management games are another form of management simulation. They usually require the participant to work, cooperatively and/or competively, with other participants. Whereas the in-basket measures a narrower range of abilities centered around administrative skills, games usually tap a broader spectrum of skills including ability to work under stress, leadership, interperconal relations, and organization and planning. Because there is such a wide variety of games in use, it is hard to

generalize about their makeup. Two games which contain some common features of many others are briefly described:

Conglomerate - Teams of participants trade shares of company stock in order to form conglomerates. To win, a company must plan and organize its activities during three fast-paced trading sessions. Behavior flexibility, interpersonal sensitivity, operation under stress and leadership are some of the evaluations made by assessors.

The Keyboard Problem - Teams of six candidates play the role of the Board of Directors of a small business concern. They buy and sell keys used on Keyboard consoles used in computer products. Given a selection of keys differing in profit margin each team must decide how to invest its money, organize its purchasing, control stock and sell its product. Assessors watch for indications of emergent leadership, organizational ability, financial acumen, quickness of thinking and operation under stress. Adaptability is observed when bonus options occur throughout the problem calling for the drastic redeployment of resources.

Hardesty & Jones (1968) found no relationship between performance in a stock market game and the OAR but a manufacturing problem was significantly correlated with both OAR and subsequent salary progress in the AT&T studies (Bray & Grant, 1966). The game seemed somewhat more predictive for the non-college samples than for the college samples. A manufacturing game also showed predictive validity at IBM (Wollowick & McNamara, 1969).

Paper and pencil tests

As one might expect, a wide variety of tests have been used in assessment centers. In no case have tests been used in a strictly

psychometric manner by setting up cutting scores and the like. Usually the tests are considered during the review period, along with all other data, to form the overall evaluation of the participant.

In the cases where we have data on the role that intellectual tests play in assessment center results, the findings are generally mixed. Since a wide variety of tests have been used with a number of different managerial positions and since cross-validation is seldom possible, it is hard to generalize about the finding or make firm statements about the applicability of specific tests. On the other hand, the results are such to warrant the recommendation to include some test of intellectual ability in the assessment center battery of assessment techniques.

Bray & Grant (1966) used three mental abilities tests in the Management Progress Study: School and College Ability Test (SCAT), Critical Thinking in the Social Science Test, and the Contemporary Affairs Test. All three of the tests correlated significantly with the staff prediction of success and with the staff rating on General Effectiveness. Correlations were found to be higher for the non-college groups, but in all cases there was much variance in the judgments not accounted for by the intellectual tests. Since the Verbal portion of the SCAT correlated highest with the staff judgments it was included in further analyses that studied the relative contributions of several assessment techniques. In the college sample the SCAT was least important (following the three situational exercises), but in the non-college sample it was second most important. When the three tests were correlated with the criterion of salary progress, the results are highly variable, but the SCAT Verbal showed a significant relationship for four of the seven groups. In one of those cases, though, the

correlation was negative. Bray & Grant (1966) give no explanation for this unusual finding. While we can see that the mental ability tests do contribute to the predictions in the Management Progress Study, it is clear that they do not tell the whole story. When the mental ability tests scores are partialed out of the staff judgment ratings, these staff judgments still correlate significantly with salary progress. This indicates that the more elaborate assessment process makes additional contributions over what is possible with only the mental tests. For the Management Progress Study, then, Bray & Grant (1966) conclude that, even though situational tests contribute more to the assessment paper-and-pencil instruments measuring mental ability should be included in the battery of techniques.

Similar results have been found in a number of other studies. Hardesty & Jones (1968) found correlations over .40 between the parts of the School and College Ability Test and the Miller Analogies Test and the Potential Rating given by the entire assessment staff. The Potential Rating was a joint decision of two professional assessors and three managers as to the individual's probable level of attainment in the company. While these correlations are significant and there are significant differences in the average test performance between high potential and not high potential individuals, the authors present graphs which show that there is great overlap of the distributions of these two groups. In other words, many individuals rated high potential got scores on the test lower than those rated not high potential.

Bray & Campbell (1968) report on the use of an assessment center to select salesmen in which the SCAT, Critical Thinking, and Abstract Reasoning tests were administered. The three tests were correlated with a set of field ratings of sales performance .25, .26, and .02

respectively.

Tests of intellectual ability have not universally proven valid in assessment center settings. Intelligence tests were used in the assessment program reported by Hinrichs (1969). The Concept Mastery Test (CMT) and the School and College Ability Test (SCAT) Numerical Part were correlated with external (relative salary standing), internal (assessment staff evaluation) and parallel (managerial potential evaluation following review of normal personnel data) criteria. There was esentially no correlation between the mental ability tests and the criteria in this study. This lack of correlation does not, of course, mean that there is no relationship between mental ability and success, and it should be noted that the study is concurrent in design. The more important information would be evidence regarding the predictive validity of the tests in relation to subsequent criteria of managerial success. Data on this last point has been reviewed in the discussion of the Bray and Grant (1966) report of the Management Progress Study.

Other studies with a predictive validity design are those of Albrecht, Glaser, and Marks (1964) and Wollowick and McNamara (1969). Albrecht, et. al. (1964) used the Problems test (a brief mental ability test) and the Watson-Glaser Critical Thinking Appraisal (a test of functional effectiveness in applied reasoning). The criteria were ratings on the following job performance dimensions: forecasting and budgeting effectiveness, sales performance, effectiveness in interpersonal reàationships, and overall effectiveness. These ratings were made by four different judge groups: district managers, regional managers, peers, and consultants. Four of the possible 32 correlations (2 tests, 4 criteria, 4 work areas) were significant, whereas 9 of 16 correlations involving the overall assessment ratings were significant.

The criterion in the predictive study of Wollowick and McNamara (1969) was the increase in managerial responsibility after the assessment procedure and it included a consideration of number of persons supervised, job complexity, and financial and skill responsibilities. Correlations of the SCAT-Total (.11) and the Otis (.07) were not significant. On the other hand, the SCAT Verbal and Quantitative correlated significantly with a composite criterion of behavioral ratings in a study at SOHIO (Carlton, 1970).

The experience with personality tests in assessment centers has been a mixture of moderate to little success. This experience reflects the general level of success with these tests in other managerial selection and placement settings. Personality tests have been used in almost all centers and continue to be used despite moderate success. Job analysis of the managers's job show that certain affective and interpersonal characteristics exist and should prove useful in assessment. These same characteristics can also be observed in other ways and probably are in the many activities of an assessment center.

In the Management Progress Study, Bray & Grant (1966) used two multi-scale tests: The Edwards Personal Preference Schedule (EPPS) and the Guilford-Martin Inventory of Factors GAMIN (GM). Only 2 of the 13 scales of EPPS and 1 of the 5 scales of the GM correlated with the overall staff prediction. When the scores were correlated with the subsequent criterion of salary progress, the results were mixed for the seven separate groups studied. The most success was found in one group in which 5 of the EPPS were significant. For the other groups the EPPS was not predictive and in all but one group the GM completely failed to predict the long range criterion.

For Hinrichs (1969) the Allport-Vernon-Lindzey Study of Values did

not correlate with the concurrent criteria. To a small extent, the Gordon Personal Preference Schedule (GPP) did correlate with the internal criterion and one scale of the same test also predicted the long range criterion in the Wollowick and McNamara (1969) study.

Carleton (1970) reports the predictive correlations of personality tests with both a global composite criterion (based on salary growth and promotions over a 4-year period) and managerial ratings of 13 scales reflecting important managerial characteristics, e.g. amount of participation, oral communication, impact, etc. The Gordon Survey of Interpersonal Values (SIV) and the Welsh Figure Preference (WFP) both showed a couple of correlations with the global criterion, but less ability to predict the rating scales, even those scales supposedly reflecting non-cognitive behaviors.

Projective Techniques

Grant, Katkovsky, and Bray (1967) report that projective techniques made contributions to the predictions of the assessment center staff and to a subsequent criterion of salary progress. After reviewing the results of three projective techniques, clinical psychologists rated the managers on nine rating scales for variables such as optimism-pessimism, general adjustment, self-confidence, leadership role, etc. These ratings were found to correlate with various components of the assessment center staff evaluations and the overall staff prediction. The variables measuring leadership and motivation proved more valuable than those measuring adjustment and other personality factors.

Grant, et. al. (1967) conclude that the projective techniques make a significant contribution to the assessment process, but very carefully warn against the over-generalization of these results:

The conditions under which these encouraging results were obtained should be noted. The projective report is a summary report based on three different instruments. This report is in no way a scoring of the projective protocols but is impressionistic in nature. Neither is the report deeply "clinical;" every effort was made to orient it to the motivations relevant to business management. Finally, because four-fifths of the reports were written by one psychologist a question could be raised regarding the replicability of the findings.

Finley (1970) reported that 13 psychological predictions based on information from two projective techniques were correlated both with the predictive ratings from the entire assessment center staff after considering all sources of information and with supervisors' ratings gathered several years later. The projectives did not correlate with supervisors' ratings in as many cases as they did with the committee's ratings, but the large number of significant correlations suggest future research may be fruitful.

Several comments are warranted with regard to this study. It should be recognized that there was contamination of both the committee ratings and the supervisory ratings from knowledge of the projective information and the supervisors' ratings were further contaminated from knowledge of the committees ratings at the end of the assessment program. Finally, there was no cross-validation of the results, although replication was present in the form of separate analyses for two different groups of managers.

Interview

Contrary to much of the literature on interviewing, it has been found by Grant & Bray (1969) that the interview can make a positive contribution to the assessment management potential. Since this study is so controversial and relevant, it warrants some detailed attention. The interviews were unstructured and conducted by psychologists as a part of the Management Progress Study at AT&T (Bray, 1964). Inter-

viewer reports were coded by judges and rated on 18 variables, including such things as Personal Impact, Oral Communications Skills, Behavior Flexibility, etc. Inter-code reliabilities were generally in the .80's with only 6 of 36 correlations (18 variables for 2 samples) below .70. Correlations of interview variables with several composite assessment variables based on all assessment techniques showed that the interview made substantial contribution. The most important interview variables seemed to be personal impact-forcefulness, oral communication skills, energy, and need advancement. Correlations of interview variables with staff predictions of whether the participant would reach middle management within 10 years revealed 22 of 35 significant relationships. Finally, 18 of 36 correlations between interview variables and a subsequent criterion of salary progress were significant. The authors clearly acknowledge that these analyses do not show the relative effectiveness of the interview in comparison with other assessment techniques. The study does establish that the interview (in this case unstructured and carried out by trained professionals) makes a reliable and valid contribution to the assessment of managerial potential.

Summary

To summarize the information in this section, I might make the following points:

(1) The most valid techniques seem to be performance exercises involving "samples" of behavior. These are to be contrasted with the more traditional paper-and-pencil tests which might be called "signs" of subsequent behavior.

(2) Note that the interview and the projective devices make valid contributions in at least one study. This is an exception to the first point just made and is in contrast to the vast

amount of generally negative literature on these techniques in previous industrial psychology studies.

(3) Few studies have treated the techniques psychometrically. Rather judgments or ratings are made after the assessors have observed the participants carry out the exercise. Thus, the validity data are usually a confounding of behavior elicited and judgments as recorded by the assessors.

(4) Numerous research and technical reports fail to report the correlations of the tests, games, and other exercises with any sort of criteria, even though these data were gathered. Practical efficiencies and theoretical advances could be made with very little additional expenditure of energy.

Validity of Overall Assessment Ratings

To address the second general area, the paper reviews the validity of overall assessment ratings (OAR) and recommendations resulting from assessment and the effectiveness of assessment centers as a management development technique. In this section, the OAR is considered the "independent variable" or the "predictor variable," the validity of which must be established by reference to concurrent and/or subsequent criteria. Table 2 summarizes the validity of the OAR as a predictor of management potential. Ideally, we would want evidence derived from longitudinal, predictive validity designs, but we shall see that concurrent studies have been carried out. The review of research will be concerned with a variety of criterion measures including administrative decisions such as salary, salary progress, and promotions, and also some more behaviorally oriented criteria such as observations of job performance by researchers and supervisors. To the extent possible, we will include evidence that shows the validity of assessment centers over and above other approaches, such as tests and supervisors evaluations of potential.

We have attempted to organize the validity research along a continuum of quality of research design. At the one end are those designs which yield the most conclusive and generalizable results. These would be long range longitudinal studies involving representative samples with little or no criterion contamination and valid criteria. At the other end are those designs which are concurrent, involve small and unrepresentative samples, and involve poor criterion likely to be contaminated by knowledge of predictor data.

Table 2 contains sufficient detail that the reader can get a pretty good understanding of these several research studies. Therefore, the body of the paper will not repeat a description of all of the detail of the research projects. The first two studies are the only two which made no operational use of the OAR. This avoids the possibility of criterion contamination through knowledge of the predictor data. The results show valid prediction of the following criteria; salary progress, management level attained, and ratings of sales performance by a specially trained field review team. In the Bray & Grant (1966) study the OAR correlated moderately high with the criterion of salary progress in four separate validation groups. For the total sample expectancy data are presented. The table shows that a significantly greater portion of those who received an OAR of "Yes, will attain third level management in 10 years" actually had progressed that far in comparison with those who received a less favorable OAR. These criteria were gathered anywhere from 5-8 years after the assessment and the results are even more clear-cut after a full 10 years (Bray, 1971; personal communication). The sales recruits study (Bray & Campbell, 1968) showed a significant correlation of OAR and ratings by field review team, but no correlation against ratings by supervisors and trainees.

The continuance of Table 2 shows some of the data regarding the unique contribution of the overall assessment process. In both studies, the OAR adds some additional ability to account for the criterion after the test data are controlled for. In most cases these are sizable increases.

Operational use of predictions; control group. In the next group of studies, the assessment results were used in the company for promotion and development purposes, therefore the possibility of criterion contamination though knowledge of the assessment evaluations is a potential changer. The other feature of the designs of this group of studies is the comparison of criterion data for a control group who were not assessed. An additional reservation with regard to this design is the possibility that any positive regard for the assessemnt center program among managers may lead to spuriously high ratings for assessed groups with little regard for actual performance.

The first two studies are ones done at AT&T involving similar designs and criteria. Criterion data were obtained for groups of managers who had been promoted to a supervisory job. Both a composite performance measure (including the man's last formal appraisal rating and a rating and ranking by the supervisor) and a composite potential measure (including the rating of potential from the last formal appraisal, a potential ranking by the supervisor, and the present level of the man) were obtained. Some of the subjects had been assessed and some had not been assessed. In the Bray & Campbell (1967) study, among the assessed group, the OAR ranged from "acceptable" to "not acceptable". The data for the performance criteria show that there is a relationship between OAR and the criteria, but that a large portion (55-63%) of the non-assessed managers who had been promoted were above average. The data on the potential criteria are more clear-cut, showing higher success rate among assessed

than non-assessed groups. Both the performance and potential criteria in the 1964 Bray study shows predictive validity.

The validation study of the assessemnt center at the Oak Ridge Gaseous Diffusion Plant of the Union Carbide Corporation (Bender, Calvert, Jaffee, 1970; Jaffee, Bender and Cavert, 1970) includes some unique features. Due to the small numbers of persons assessed at the time of the study, it was not possible to compare on-the-job performance of persons assessed at various levels of acceptability. Therefore, the design included a control group of 13 subjects promoted immediately prior to the initiation of the assessment program and an experimental group of 13 subjects were assessed. Analyses showed that the two groups were comparable on a number of important variables, but that the experimental group was younger, had less company service, and had been in supervision a shorter period. Criteria information were collected in two areas (1) objective data of performanace of the subject's work group, i.e. absences, grievances, or visits to the infirmary, (2) interview data from the supervisors and randomly selected subordinates regarding the adequacy of the subject as a supervisor.

No differences in the objective criterion were found between the experimental and control groups. The study was done about one year after assessment and thus "considering the short time in a supervisory capacity for most of the experimental group it is not at all surprising that these data did not reveal any meaningful differences (Bender, et. al., 1970, p. 32)."

The interview data showed that more positive comments and fewer negative comments were made about the performance of the experimental group than about the control group. This pattern held for both the superiors and the subordinates (except that the subordinates made more negative comments regarding assessed subjects). The authors warn that "These results must certainly be considered carefully and no conclusions of a definite nature

may be drawn, but they are certainly indicative of a trend (Bender, et. al., 1970, p. 33)." A major limitation of the study of course is the small sample size and the potential positive halo surrounding the assessed group.

The studies we have reviewed up to this point have been generally supportive of the predictive validity of the assessment center techniques. In my critique of these studies we have noted the design flaws which may limit the usefulness of the results and their generality. The only published study (Bullar, 1969) which is wholly negative involves the assessment center at Caterpillar Tractor Company. Bullard (1969) reports a study of (1) the relationship between assessment center predictions and two measures of future managerial behavior and (2) a comparison of the assessment center and a traditional method of selection and placement.

Caterpillar's assessment center is very similar to the ones who have reviewed thus far. Included in the assessment battery are a number of situational exercises and leaderless group discussions. Following the two-day evaluation period, the assessment staff spent the remainder of the week consolidating their observations rating the individual's performance on a 25-item Supervisory Qualification Rating Form (SQRF) and deriving an overall rating for each candidate: more than acceptable, acceptable, not ready now, and low in managerial skill. The SQRF is used as one of the criterion measures, being filled out subsequently by the general foreman on the job. The SQRF includes 25 behavioral characteristics considered important managerial skills and traits, e.g. company attitude, importance of work, inner work standards, motivation, drive, etc. A brief description of each trait is given on the form followed by a five-point rating scale from low to high. No additional descriptors or qualifiers are given for the scales points. The form may suffer from lack of clarity and objectivity. No data on its reliability are provided. The second criterion was

the Personal Description Check List (PDCL), a list of 108 specific on-the-job behaviors, e.g. rarely takes chances, cooperates when requested, is too shy to be a leader, etc. This form was not used during the assessment center so no direct comparison can be made for this instrument.

Criterion data were gathered for 37 experimental (assessed) subjects comprised of 4 groups from different plant locations and for 27 control subjects from the same locations. The control subjects had been placed on their job by the traditional method. Bullard (1969) states "Absolute control for age, time on supervision, and tenure with the Company was difficult in view of the small N (p5)." The report is not clear as to when the experimental and control supervisors were placed on their jobs. If the control subjects were promoted prior to the institution of the assessment center several problems could arise. First, the controls would have been on the job for a longer period at the time of criterion measurement; it would be likely that their job performance would be greater. Second, it may be that the experimental subjects were drawn from a very different pool of potential supervisors than the controls. The controls may be the "cream of the crop" selected at one point in time and then the experimental group was selected from a pool with generally lower abilities. This is pure speculation, for the Bullard (1969) report does not provide sufficient information.

With the question of comparability of subjects in mind we can examine the mean criterion scores for experimental and control groups in Table 2. On both the PDCL and the SDRF, the control group scored higher than the experimental group. Analyses for the subgroups at the 4 locations follow the same pattern. Bullard (1968) interprets these results "as evidence that if superiority is to be assigned to either group on the basis of mean performance, then it must be assigned to those supervisors which constitute the control or 'traditional method' group (p.12-13)."

Operational use of predictions; no control group. The last group of studies are those involving no control group and ones in which assessment data were used in the company. The data are generally supportinve, but the possibility of criterion contamination limit the generalizations from the fiudings.

The studies of IBM (Hinrichs, 1969; Wollowick & McNamara, 1969) provide evidence of the predictive validity of assessment center judgments and also show in what ways the situational exercises contribute to the evaluations. Wollowick and McNamara (1969) report a study of 94 lower and middle managers who were followed up 3 years after assessment. The criterion was increase in managerial responsibility based on a complex index including number of persons supervised, complexity of job, financial responsibliity and skill requirements of the job. There were 12 increments in this index and the suthors believe it is a better measure of advancement than simply level in the organization. Several results from the suudy are presented in Table 2 and its appendix. The first correlation (r=.37) indicates that the overall assessment rating of management potential is a fairly accura e predictor of progress. But the best combination of psychometric tests correlated .45 with the criterion, suggesting that some predictive efficiency is lost when the judgmental process works on all the assessment data. This higher correlation with the multiple regression equation may result from statistical artifacts in compitalizing or chance errors. All results should, of course, be cross-validated. Of even more significance is the jump from .45 to .62 when then the tests and exercises snd included in the multiple regression analysis. The percentage of criterion variance accounted for increases from 20% to 36%.

Wollowick and McNamara (1969) put forth two tentative conclusions based on these results. First, the situational exercises contribute a significant

amount to the assessment of important characteristics related to management success (R increases from .45 to .62). Second, the mechanical combination of separate sources of data gathered in the assessment center may improve predictive effectiveness over and above the subjectively determined overall rating. The first conclusion is similar to other results we cited in earlier sections; the second conclusion fits in with the general argument in favor of mechanical and actuarial prediction methods. No other study or assessment centers makes a comparison such as Wollowick and McNamara, but their study suggests strongly that some type of systematic weighting of inputs may improve predictions of progress.

Wollowick and McNamara warn that their results should be cross-validated and this is always sound advice in multiple regression studies with a large number of variables and a small number of subjects. In lieu of actual cross-validation, it is possible to get some idea of the results we might expect. "Shrinkage" formulas, although not completely satisfactory (Horst, 1966) give an estimate on cross-validation to a new sample (Blum & Naylor, 1968). When this type of analysis is done with the data in Wollowick and McNamara (1969) we see only a slight reduction in multiple correlations and percentages of criterion variance accounted for. Another cause for placing some confidence on the results of this study is the generally acceptable reliabilities reported for the situational exercises (Greenwood & McNamara, 1967). Ratings and rankings of the leaderless group discussion, task-force committee discussion, and the manufacturing game were shown to have inter-rater reliabilities predominantly in .70's and .80's. For one group of subjects the reliabilities were more typically in .60's. Of further note, were the comparisons between professionals and non-professionals. These correlations were somewhat lower (in a few cases near zero) but the N's were small in all cases. In summary "The results of this study suggest

that reasonable rater reliability can be obtained in situational tests commonly used to assess . . . potential for advancement (p. 105, Greenwood & McNamara, 1967)."

Hinrichs (1969) provides further data on the IBM assessment programs in a study of 47 marketing recruits being considered for first level management. The study is unique in several regards and requires some close attentia For our purposes in this section, we can examine three variables in the study: (1) the overall evaluation of management potential resulting from deliberations of results of situational exercises and tests (internal criterion), (2) a measure of relative salary standing (external criterion), and (3) managers' evaluation of potential (parallel criterion). Each of these criteria are examined more closely below. Of particular interest to us now are the data in Table 3. Based on these and other data, Hinrichs (1969) concludes:

> The data suggest that traditional approaches to the assessment of management potential in the form of careful evaluation of personnel records and employment history (our parallel criteria) can perhaps provide much the same information which evolves from the lengthy and expensive 2-day assessment programs . . . (p. 43).

We suggest that the degree of overlap between the two assessment methods (which in fact is not highly substantial) is not adequate evidence of the effectiveness of either of them for assessing management potential. Hinrichs acknowledges that the data are not adequate for judging validity and we agree. The author does tend to cast the overall assessment evaluation in terms of a "criterion" for comparison with other "criteria." This places high value on the assessment center approach.. We would prefer to view the assessment center as a predictor and question its value in relation to harder criteria. The use of the external criterion of relative salary is not entirely adequate for this purpose. The purpose of the program was the prediction of management potential - a variable requiring longitudinal measurement. The salary criterion was a measure of current value based on the relationship

of the individual's salary in comparison with that of his peers.

Table 3

Intercorrelations Among Criteria

Variable	1	2	3	$r_{13\text{-}2}$
1. Relative salary standing (external criterion)	--	.10	.37	.37
2. Manager's potential evaluation (paralled criterion)		--	.46	
3. Assessment program evaluation (internal criterion)			--	

In the past five years, Wickes has used two assessment programs for retail lumber yard employees (Ague, 1971). The four-day program for assistant managers is primarily developmental and enables Wickes to determine when they are ready for promotion. In five years 184 assistant managers have been assessed and 126 have been promoted to manager. The annual turnover rate of assistant managers dropped form 30% prior to the program to 8% for assessed assistant managers. The Wickes Hourly Assessment Program has processed 240 hourly employees in the past five years. One hundred and twenty nine have been promoted to their management training program. The annual turnover for these 128 hourly employees, after initial management assignments is about 3% (Ague, 1971). No comparative data for the hourly turnover figure were presented.

Turnover statistics reported by Wickes are a unique criterion which other companies might also use. These figures at Wickes represent very low turnover rates in the years after the program was initiated. From the information available, it is impossible to tell whether the reduction in turnover is due to the assessment program or to some other changes in the company or in the economy as a whole. If the company launched a broad change in managerial philosophy or the company was experiencing rapid growth and financial success, there are any number of potentially valid explanations

of reduced turnover. Certainly more information is needed, but the results are very supportive and favorable for the impact of the assessment process.

Although the assessment center approach has been used most frequently in industrial settings, it has also been applied in government organizations. A few of these include the Tennessee Valley Authority (Eppenbach, 1971), Public Service Commission of Canada (Grant & Slivinski, 1970; Slivinski & Grant, 1970), and the U.S. Internal Revenue Service (DiCostanzo & Andretta, 1970; Eeny, Meeny, miney . . . Supervisor?, 1970). Relatively little data exists on the validity of assessment centers in these types of settings. The IRS has carried out a small scale evaluation project (DiCostanzo & Andretta, 1970). Eleven offices were involved and 142 persons were assessed. In discussing the results in comparison with traditional evaluative information, the authors note that "there was no correlation between relative strengths and weaknesses shown on (supervisory) evaluations and the corresponding areas in the assessment center reports . . . feedback from local management officials who used the reports indicated that in many cases they probed deeper with the supervisor and satisfied themselves that the Assessment Center findings were correct (p. 14, DiCostanzo & Andretta, 1970)

The problems of obtaining good criterion data are apparent in the IRS situation as we have observed in other companies. The problem of contamination is magnified by an orientation program run for managers. The IRS ran three two-day assessment center workshops, at which time over 75 higher level managers learned of the history and theory of assessment centers, the detail of a program, and actually went through a miniature assessment program. The important thing is that all these individuals were persons who would have access to and who would use the assessment center results (DiCostanzo & Andretta, 1970). It is quite possible that a favorable set might be de-

veloped toward any supervisor who is recommended by the assessment process, thus precluding the objective assessment of on-the-job performance.

The TVA initiated a pilot program in an assessment center for managers in the divisions of engineering design and construction, water control, forestry, and navigation (Epenbach, 1971). The standard types of exercises were used and the evaluators (higher level managers) met to compile their ratings of performance. Independent judgments were made by the evaluators on each candidate for 15 managerial skills. Erpenbach (1971) concludes:

> The MSAP (Management Skills Appraisal Program) was judged to have been highly successful by the experimenters. They found a high degree of relationship between the appraisers' ratings and the ratings of the appraised employees' supervisors. Some at TVA, however, feel that the supervisors' ratings alone are adequate since they are best qualified to assess their employees' potential. Because of budget limitations the program was discontinued, and there are no current plans to continue it (. 2).

Several interesting inferences can be drawn from the brief amount of information available.

The assessment programs of Standard Oil (Ohio) have been studied and reviewed quite thoroughly by their corporate staff (Finkle and Jones, 1970; Finley, 1970; Carleton, 1970) and outside researchers (Donaldson, 1969; Thomson, 1969, 1970). These studies have provided very thorough analyses of the internal and external validity of the assessment data. Of particular interest is the relationship between assessment ratings and on-the-job ratings for the same 13 behavioral traits. Three different sources (Carleton, 1970; Finley, 1970; Thompson, 1970) provide data relevant to the predictive validity of assessment trait ratings for trait rating of the job some years in the future. It can be concluded that assessors are able to validly predict behavior on several behavioral dimensions as a result of assessment center activities. Most of correlations are low to moderate, but the relationships of ratings of Potential are consistently high (i.e. in the .65 range).

Two traits (Orientation to detail and Relationship to Authority) account for 6 of the 7 non-significant correlations. This may be partially explained by the hypothesized curvilinear relationship between these traits and success. Carleton (1970) states that for these traits "intermediate ratings were considered more favorable than either high or low. (p. 565)"

The criteria in these studies could certainly be subject to bias and contamination but the preas to distort the results is probalby minimal in comparison with the problems inherent in administrative criteria such as promotiona and salary data used so frequently in other studies we shall review below. The rating acales themselves used to gather the criterion ratings seem entirely adequate although we are not told how they were developed. Each of the rating instrumenta is labled, provides a definition of the trait, and includea a behavioral description of each of the five scale points (Thompson, 1969).

Several questions might be aaked about these data. Since the correlations are only moderately high, it is poasible thac they are a result only of autocorrelation or method communality. Why aren't they higher? Thompson (1970) has provided data relevant co theae questiona and many more in a multitrait-multimethod analysis of ratings. The matrix of correlations between 13 traita and 3 sets raters (management assessors, psychologist assessors, and supervisors) were examined for convergent and discriminant validity. High levels of convergent validity (median r = .85) and discriminant validity were found for the two aaaessor groups on the 13 traits, but the supervisors ratings showed lack of discriminant validity. The supervisors ratings showed a general halo effect with little differentiation among the 13 traits. Thomson (1970) suggests that "the criterion raters were unable to arrive at a precise and common understanding of the meaning of the different scales, and responded to some generalized notion of the goodness or badness of the assessee (p. 501)."

Summary and Conclusions. A review of the data in the second part of the paper suggest the following points:

(1) The overwhelming general finding is that the OAR is related to a variety of criteria of subsequent performance and potential in managerial positions.

(2) Serious questions regarding the quality of research designs and problems with criterion contamination in some studies limit the generality of the findings.

(3) The overall assessment evaluation adds a unique contribution to assessment over and above the information which can be obtained from objective tests.

Final Summary

The published and private literature on assessment centers is generally supportive of the predictive validity of evaluations of management potential during the assessment process. Major design flaws in some studies make conclusions difficult. Of over 30 studies reviewed involving 21 companies, only one involved clearly no validity (Caterpillar) and one included strong reservations (Hinrichs, IBM). Where comparisons are made with other methods of assessing management potential, assessment centers make unique and substantial contributions.

Virtually no objective data has been published or is available from sponsoring organizations which support the claim of the effectiveness of assessment centers as a management development technique. While admittedly it is difficult to measure some of the subtle outcomes that may be desired the burden of proof must fall on those who advocate the use of assessment centers for such purposes.

Table 1

Validity of Individual Assessment Techniques

	Internal Validity	External Validity
In-basket		
Bray & Grant, 1966	College 55	College groups: 27,-.01, 03,28
	Non college 51	Non college groups: 44, 22, -19
Wollowick & McNamara, 1969		32
Leaderless Group Discussion		
Bray & Grant, 1966	College 60	College groups: 30,50,26, 38
	Non college 38	Non college groups: 33, 28,10
Wollowick & McNamara, 1969		25
Management Games		
Bray & Grant, 1966	College 41	College groups: 15,41,14, -01
	Non college 42	Non college groups: 37, 50,29
Hardesty & Jones, 1968	-05	
Wollowick & McNamara, 1969		.28
Tests		
SCAT:		
Bray & Grant, 1966		
College	V36, Q06, T27	V36,35,51,30,19,-44,14
Non college	v44, Q29, T41	Q23,44,-04,19,09,-10,-28 T38,45,32,28,18,-30,-03
Wollowick & McNamara 1969		T11
Bray & Campbell, 1968		25
Hinrichs, 1969	T03,03,-13	
Hardesty & Jones, 1968	V48, Q42, T53	
Carlton, 1970		V29, Q36

Table 1 (cont.)

	Internal Validity	External Validity
(Note: For the following personality testa, data are number of scales with significant correlations.)		
EPPS:		
Bray & Grant, 1966	2	5,3,1,1,1,0,0
Gordon Tests:		
Hardesty & Jones, 1968	0	
Hinrichs, 1969	2,0,2	
Wollowick, 1969		1,1
Carleton, 1970		1,1
Study of Values:		
Hinrichs, 1969	0,0,0	
Projective Techniques		
Grant, et. al., 1967		For several projective
College	35	variablea -01,10,11,-06,
Non college	40	16,24,-35,-25,26,12,19, 21,-07,17,29,120,-23,30
Finley, 1970		
Sample I:	11 of 13 v's significant	3 of 13 v's significant
Sample II:	13 of 13 v's significant	7 of 13 v's significant
Interview		
Grant & Bray, 1969		
College	14 of 18 v's significant	9 of 18 v's significant
Non college	8 of 18 v's significant	9 of 18 v's significant

Table 2 Validity of Overall Assessment Ratings (OAR)

No operational use of predictions

Reference	Subjects	Criteria	Validity
Bray & Grant, 1966, AT&T	123 college & 144 non-college non-supervisory	Salary Progress	.48, .57, .51, .52 for 4 samples
		Management level attained in 1965 (5-8 years post assessment)	

OAR	Level attained 3-4	2	1
-Yes, will attain 3rd level.	42%	54%	4%
-No, will not; or ?	7	51	42

Bray & Campbell, 1968, AT&T	78 sales recruits	Ratings by trainers, supervisors, & field teams of attainment of performance standards.

OAR	N	% meeting field review stnds
More than acceptable	9	100%
Acceptable	32	60%
Less than acceptable	16	40%
Unacceptable	21	10%

Field review team	.51
Supervisors	n.s.
Trainers	n.s.

Operational use of predictions; control groups

Campbell & Bray, 1967, AT&T	223 assessed & 283 non-assessed non-supervisory	Performance Ratings & Rankings Potential Ratings & Rankings

OA OAR	% above averate perf.	% above average pot.
Acceptable	68	50
Questionable	65	40
Not acceptable	46	31
Men Not Assessed		
Post-Assess. Promotions	63	19
Pre-Assess. Promotions	55	28

Bray, 1964 AT&T	40 assessed & 40 non-assessed vocational employees	Performance Ratings Potential Ratings	Groups Assessed Non-assessed	% better than satisf. 63 35	% showed potential for advancement 68 35
Bender, et al, 1970 Jaffee, et al, 1970 Union Carbide	13 assessed & 13 non-assessed lower managers	(1) Objective data on work group (2) Interviews with superiors & subordinates of Ss.	(1) no differences in experimental and control subjects (2) more positive comments and fewer negative comments regarding assessed supervisors.		
Bullard, 1969 Caterpillar	37 assessdd & 27 non-assessed non-supervisory	On-the-job behavior ratings by supervisors on two different forms.		Assessed	Non-assessed
			Behavior check list $\bar{X}$	181	195
			Perform. Rating Form $\bar{X}$	84	89
			(N.B. Controls scored significantly higher)		

Operational use of predictions; no control group

Wollowick & McNamara, 1969 IBM	94 lower and middle managers	Increase in managerial responsibility over 3 years	.37
Hinrichs, 1969 IBM	47 marketing recruits	(1) Relative salary standing (2) Manager's evaluation of potential	(1) .37 (2) .46
Agne, 1971 Wickes	184 asst. managers	Turnover rate of managers	Decline from 30% to 8%
	250 hourly employees	Turnover rate of asst. mgrs.	Turnover now 3%

eton, 1970
ey, 1970
Thomson, 1964
SOHIO

Relationships of Assessment center ratings and supervisory ratings on 13 draits in three different studies.

	Finley		Carleton	Thomson	
	Assessment Committee vs. Supervisor		Assessment Committee vs. Supervisor	Psychologists vs. Supervisors	Managers vs. Supervisors
Scale	Sample 1 (N-109)	Sample 2 (N=119)	(N=122)	(N=71)	
Amount of participation	30*	62*	35*	58*	65*
Oral communication	23*	39*	26*	44*	37*
Personal acceptability	29*	31*	31*	33*	37*
Impact	20*	36*	24*	50*	47*
Quality of participation	22*	25*	21*	31*	38*
Personal breadth	28*	42*	29*	55*	52*
Orientation to detail	00	19*	01	19	14
Self-direction	20*	26*	24*	42*	35*
Relationship with Authority	15*	31*	12	35*	37*
Originality	29*	43*	28*	45*	45*
Understanding of people	17	40*	20	35*	39*
Drive	26*	21*	29*	12	29
Potential	65*	63*	65*	64*	64*
Median	23	36	26	42	38

Table 2 con't.

Unique Contribution of OAR

Bray & Grant, 1966	Correlation of OAR & salary progress	.48 .57 .51 .52
	Partial correlation with ability test constant	.32 .39 .29 .42
Bray & Campbell, 1968	Correlation of OAR & field ratings	.51
	R of best combination of 4 tests	.33
Wollowick & McNamara, 1969	Correlation of OAR & criterion	.37
	Paper & pencil tests	.45
	Tests, characteristic, & exercises	.62
Hinrichs, 1969	Correlation of OAR & salary	.37
	with manager's evaluation partialed out	.37

Assessment Centers
Reading Materials
compiled by
George C. Thornton III
Colorado State University

Acker, S.R. & Perlson, M.R., Can We Sharpen Our Management of Human Resources?, Corporate Personnel Department, Olin Corporation, June 1970.

Albrecht, P.A., Glaser, E.M., & Marks, J., Validation of a multiple assessment procedure for managerial personnel, Journal of Applied Psychology, 1964, 43, 351-359.

Albrook, R.C., How to Spot Executives Early, Fortune, July, 1968, 78 (2), 106-111.

Bender, J.M., Calvert, O.L., & Jaffee, C.L., Report on Supervision Selection Program--Oak Ridge Gaseous Diffusion Plant, Union Carbide Corporation, Nuclear Division, April 17, 1970 (K1789).

Bray, D.W., The assessment center method of appraising management potential. In J.W. Blood (Ed.), The Personnel Job in a Changing World, New York: American Management Association, 1964.

Bray, D.W., The management progress study, American Psychologist, 1964, 19, 419-420.

Bray, D.W., In-Basket as Used in the Assessment Centers, unpublished paper (excerpt from a talk by D.W. Bray, New York Workshop on Assessment Centers, 1969).

Bray, D.W., & Campbell, R.J., Selection of salesmen by means of an assessment center. Journal of Applied Psychology, 1968, 52, 36-41.

Bray, D.W., Assessment Centers, unpublished paper given at New York Workshop on Assessment Centers, 1969.

Bray, D.W., & Grant, D.L., The assessment center in the measurement of potential for business management, Psychological Monographs, 1966, 80 (17, Whole No. 625).

Bullard, J.F., An Evaluation of the Assessment Center Approach to Selecting Supervisors, Mimeo report, Caterpillar Tractor Co., May 1969.

Byham, W.C., The Uses of Assessment Centers, (A summary of material presented at the Spring 1969 meeting of the Executive Study Conference and Material compiled in preparation for the conference), unpublished paper, 1969.

Byham, W.C., Assessment Center for Spotting Future Managers, Harvard Business Review, 1970, 48 (4), 150-160, plus appendix.

Byham, W.C., & Pentecost, Regine, The Assessment Center: Identifying Tomorrow's Managers, Personnel Magazine, September-October, 1970.

Byham, W.C., & Thornton, G.C., III, Assessment Centers: A new aid in management selection, Studies in Personnel Psychology, 1970, 2 (2), 21-35.

Carleton, F.O., Relationships between follow-up evaluations and information developed in a management assessment center. Paper presented at the meetings of the American Psychological Association Convention, Miami Beach, Florida, 1970.

Cronbach, L.J., Essentials of Psychological Testing, (3rd Ed.), Harper & Row, 1970.

DiCostanzo, F., & Andretta, T., The supervisory assessment center in the Internal Revenue Service, Training and Development Journal, 1970, 12-15.

Dodd, W.E., & Kraut, A.I., Will management assessment centers insure selection of the same old types Paper presented at the meetings of the American Psychologica Association Convention, Miami Beach, Florida, 1970.

Dodd, W.E., & Kraut, A.I., The prediction of management assessment center performance from earlier measures of personality and sales training performance: a preliminary report, Corporate Personnel Research, April 1970.

Dodd, W.E., Two factors related to career progress in IBM: A summary of longitudinal studies in the identification of management potential.

Donaldson, R.J., Validation of the internal characteristics of an industrial assessment center program using the multitrate-multimethod matrix approach. Unpublished dissertation. Case Western Reserve University, January 1969.

Eney, meeny, miney...supervisor? Industry Week, 1970, November 16, 28032.

Finkle, R.D., & Jones, W.S., Assessing corporate talent; A key to managerial manpower planning, N.Y.: Wiley-Interscience, 1970.

Finley, R.M., Jr., An evaluation of behavior predictions from projective tests give in a management assessment assessment center. Paper presented at the meetings of the American Psychological Association Convention, Miami Beach, Florida, 197

General Electric, Talent Development Program (Supplement for Potential Staff Members) Mimeo Report, 1969.

Gordon, L.V., Clinical, psychometric, and work-sample approaches in the prediction of success in Peace Corps Training, Journal of Applied Psychology, 1967, 51, 111-119.

Grant, D.L., & Bray, D.W., Contributions of the interview to assessment of management potential, Journal of Applied Psychology, 1969, 53, 24-34.

Grant, D.L., Katkovsky, W., & Bray, D.W., Contributions of projective techniques to assessment of management potential, Journal of Applied Psychology, 1967, 51, 226-232.

Grant, K.W., & Slivinski, L.W., The Assessment Centre Approach: A Literature Review and Implications for the Federal Government. Personnel Assessment and Research Division, Public Service Commission of Canada, March, 1970.

Greenwood, J.M., & McNamara, W.J., Interrater reliability in situational tests, Journal of Applied Psychology, 1967, 31, 101-106.

Greenwood, J.M., & McNamara, W.J., Leadership styles of structure and consideration and managerial success, Personnel Psychology, 1968, 21, 85-98.

Guion, R.M., Personnel Testing, New York: McGraw-Hill, 1965.

Hardesty, D.L., & Jones, W.S., Characteristics of judged high potential management personnel -- The operations of an industrial assessment center, Personnel Psychology, 1968, 21, 85-98.

Hinrichs, J.R., Comparison of "Real Life" assessments of management potenaial with situational exercises, paper-and-pencil ability tests, and personality inventories, Journal of Applied Psychology, 1969, 53, 425-433.

Jaffee, C.L., Assessment Centers Help Find Management Potential, Bell Telephone Magazine, 1965, 44 (3), 18-24.

Jaffee, C.L., "Managerial assessment: Professional or managerial prerogative," Personnel Journal, 3, 1966.

Jaffee, C., Bender, J., & Calvert, D.L., The assessment center technique; a validation study, Management of Personnel Quarterly, Fall, 1970, 9-14.

Korman, A.K., The prediction of managerial performance, Personnel, 1969, March/April.

OSS Assessment Staff, Assessment of Men, New York: Rinehart, 1948.

Schaffer, A.J., Information about assessment center for ES&D Program Finalists. Memorandum from Director, Personnel Division, National Office, Internal Revenue Service, September 16, 1970.

Slivinski, R.W., & Grant, K.W., Assessment Centre Proposal, Personnel Assessment and Research Division, Public Service Commission of Canada, January, 1970.

Stern, G.G., Stein, M.I., & Bloom, B.S., Methods in Personality Assessment, Glencoe, Ill., Free Press, 1956.

Story, W., Assessment Centers, unpublished paper given at New York Workshop on Assessment Centers, 1969.

Taft, R., Multiple methods of personality assessment, Psychological Bulletin, 1959, 56, 333-352.

Tennessee Valley Authority, TVA's Experiment in the Assessment of Managerial Potential, undated.

Thomson, H.A., Internal and external validation of an industrial assessment program. Unpublished doctoral dissertation. Case Western Reserve University, January, 1969.

Thomson, H.A., Comparison of predictor and in criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology, 1970, 54, 496-502.

Wollowick, H.B., & McNamara, W.J., Relationship of the components of an assessment center to management success, Journal of Applied Psychology, 1969, 53, 348-352.